基于时空流特征融合的俯视视角下奶牛跛行自动检测方法

代 昕[1]，王军号[1]，张 翼[1,2]，王鑫杰[1]，李晏兴[1]，戴百生[1*]，沈维政[1*]

(1. 东北农业大学 电气与信息学院，黑龙江哈尔滨 150030，中国；2. 黑龙江东方学院 信息工程学院，黑龙江哈尔滨 150086，中国)

摘 要：［目的/意义］奶牛跛行检测是规模化奶牛养殖过程中亟待解决的重要问题，现有方法的检测视角主要以侧视为主。然而，侧视视角存在着难以消除的遮挡问题。本研究主要解决侧视视角下存在的遮挡问题。［方法］提出一种基于时空流特征融合的俯视视角下奶牛跛行检测方法。首先，通过分析深度视频流中跛行奶牛在运动过程中的位姿变化，构建空间流特征图像序列。通过分析跛行奶牛行走时躯体前进和左右摇摆的瞬时速度，利用光流捕获奶牛运动的瞬时速度，构建时间流特征图像序列。将空间流与时间流特征图像组合构建时空流融合特征图像序列。其次，利用卷积块注意力模块（Convolutional Block Attention Module, CBAM）改进PP-TSMv2（PaddlePaddle-Temporal Shift Module v2）视频动作分类网络，构建奶牛跛行检测模型Cow-TSM（Cow-Temporal Shift Module）。最后，分别在不同输入模态、不同注意力机制、不同视频动作分类网络和现有方法4个方面对比，进行奶牛跛行实验，以探究所提出方法的优劣性。［结果和讨论］共采集处理了180段奶牛图像序列数据，跛行奶牛与非跛行奶牛视频段数比例为1∶1，所提出模型识别精度达到88.7%，模型大小为22 M，离线推理时间为0.046 s。与主流视频动作分类模型TSM、PP-TSM、PP-TSMv2、SlowFast和TimesFormer模型相比，综合表现最好。同时，以时空流融合特征图像作为输入时，识别精度分别比单时间模态与单空间模态分别提升12%与4.1%，证明本研究中模态融合的有效性。通过与通道注意力（Squeeze-and-Excitation, SE）、卷积核注意力（Selective Kernel, SK）、坐标注意力（Coordinate Attention, CA）与CBAM不同注意力机制进行消融实验，证明利用CBAM注意力机制构建奶牛跛行检测模型效果最佳。最后，与现有跛行检测方法进行对比，所提出的方法同时具有较好的性能和实用性。［结论］本研究能够避免侧视视角下检测跛行奶牛时出现的遮挡问题，对于减少奶牛跛行发生率、提高牧场经济效益具有重要意义，符合牧场规模化建设的需求。

关键词：奶牛跛行检测；时空融合；视频动作分类；深度图像；注意力机制；TSM

中图分类号：TP391 **文献标志码：**A **文章编号：**SA202405025

引用格式：代昕, 王军号, 张翼, 王鑫杰, 李晏兴, 戴百生, 沈维政. 基于时空流特征融合的俯视视角下奶牛跛行自动检测方法[J]. 智慧农业(中英文), 2024, 6(4): 18-28. DOI：10.12133/j.smartag.SA202405025

DAI Xin, WANG Junhao, ZHANG Yi, WANG Xinjie, LI Yanxing, DAI Baisheng, SHEN Weizheng. Automatic Detection Method of Dairy Cow Lameness from Top-view Based on the Fusion of Spatiotemporal Stream Features[J]. Smart Agriculture, 2024, 6(4): 18-28. DOI：10.12133/j.smartag.SA202405025 (in Chinese with English abstract)

0 引 言

在推进奶牛的智能化和现代化养殖过程中，奶牛跛行发病率的提高成为了阻碍奶牛健康生长和产奶量提高的重要原因之一[1,2]，甚至被认为是影响奶牛动物福利进而影响牧场生产力的最严重问题之一[3]。当奶牛出现跛行病情时，奶牛会因为剧烈疼痛导致行走时蹄子着地困难，影响其正常行走、采食和挤奶，最终导致产奶水平和繁殖能力下降[4]，造成青年奶牛的过早淘汰。因此，及时发现并治疗奶牛跛行可以最大限度地减轻奶牛身体疼痛，减少牧场经济损失[5,6]。与人工观察相比，基于计算机视觉的奶牛跛行检测方法可长时间在非结构化环境

收稿日期：2024-05-31

基金项目：国家自然科学基金项目（32072788）；黑龙江省重点研发计划（2022ZX01A24）；国家重点研发计划（2023YFD2000700）；黑龙江东方学院科研平台支撑项目（PTZCXM2404）

作者简介：代 昕，研究方向为智慧畜牧和智能视觉感知。E-mail：daixin@neau.edu.cn

*通信作者：1. 戴百生，博士，副教授，研究方向为智慧畜牧和智能视觉感知。E-mail：bsdai@neau.edu.cn；2. 沈维政，博士，教授，研究方向为智慧畜牧和数字农业。E-mail：wzshen@neau.edu.cn

中工作，并且工作效率高，能够有效降低人工成本，这也是未来奶牛跛行检测发展的主趋势[7]。目前，通过计算机视觉的奶牛跛行检测方法从检测视角上主要分为两大类。

第1类为侧视视角。通常是定位奶牛蹄子、背部脊柱、头颈和头部等关键区域进行单一特征或联动特征融合来实现奶牛跛行检测[8-10]。Wu等[11]提出一种基于YOLO（You Only Look Once）v3深度学习算法和相对步长特征向量的奶牛跛行检测技术。首先根据YOLOv3网络定位奶牛的四肢，将前后肢的质心距离变化提取成特征向量，再输送给长短期记忆网络（Long Short-Term Memory, LSTM）网络进行跛行预测。由于跛行奶牛需要弓背来负担发生跛行病情时行走过程中的疼痛，因此跛行奶牛的背部弯曲程度往往偏大。Jiang等[12]提出了一种基于深度学习方法的计算背部曲率的奶牛跛行识别技术。首先通过目标检测定位出奶牛背部区域，然后通过帧间差分法提取出去除背景的奶牛背部脊柱区域，通过三点圆法计算出奶牛背部的曲率并作为特征值，输送给双向长短期记忆（Bidirectional Long Short-Term Memory, BiLSTM）网络进行训练，得到跛行与非跛行奶牛的二分类结果，在567段视频上的分类精度达到96.61%。跛行奶牛行走时蹄部往往由于疼痛承重能力下降而导致步态不规律性。Kang等[13]提出了一种奶牛跛行检测方法，通过降维的基于牛腿位置的时空图像，保留步态信息，使用DenseNet算法根据时空图像进行跛行分类，精度达到了98.5%。Zheng等[8]提出一种孪生注意力模型来实现奶牛腿部自动跟踪，通过注意力机制预测后续帧的牛腿的位置，并利用牛腿坐标计算相对步长，利用支持向量机（Support Vector Machine, SVM）模型实现奶牛跛行分类。Li等[9]提出了一种利用微小运动特征的时空聚合网络，通过设计的模块捕捉奶牛运动时的微小运动特征和时空特征来进行奶牛早期跛行的识别。针对奶牛跛行的步态不对称性，Li等[14]提出了一种基于RGB、光流和骨骼等多种特征的奶牛跛足检测方法，根据不同的输入将网络分为3个分支：对于分支1和分支3，使用卷积神经网络（Convolutional Neural Networks, CNN）根据输入图像和光流预测跛行；对于分支2，使用时空图卷积网络用于根据奶牛的骨骼预测跛足；最后调整权重，融合这3个分支的预测分数，最佳准确度达到了97.2%。然而，奶牛在牧场中的运动往往是成群结队地行走，当多头奶牛并排行走时，远离相机一侧的奶牛由于被其他奶牛或者栏杆遮挡从而导致相机无法有效地捕获相应的图像。因此，在牧场环境下侧视视角检测跛行奶牛时出现的遮挡问题，是阻碍检测方法应用的主要原因之一。

第2类为俯视视角。通过重建背部提取奶牛脊柱或单模态运动信息来进行跛行的识别。Abdul等[15]提出了一种俯视视角下的基于3D相机的奶牛步态特征跛行检测技术；定位奶牛后肢两处的髋关节和脊柱中心点，分别将髋关节的深度值与脊柱中心点深度值作差，提取深度差值运动曲线进行正弦拟合并进行希尔伯特变换，通过不同曲线的相位差来表征不同跛行程度的奶牛；在22头奶牛的数据集上达到了95.7%的精度。Arazo等[16]提出了一种基于RGB和深度视频分割增强奶牛跛行检测的技术；首先使用带有ResNeXT网络的特征金字塔（Feature Pyramid Networks, FPN）网络作为分割模型，然后使用SlowFast视频分类模型直接处理输入视频，再将输出特征输给分类器模型去实现跛行与非跛行的二分类。然而，现有的基于俯视视角下的奶牛跛行检测研究较少，主要通过单一的模态特征与姿态特征提取的方法进行检测，但俯视视角下的奶牛跛行特征并不明显，单一模态很难完整表达奶牛运动信息。

本研究利用深度图像研究俯视视角下的奶牛跛行检测方法。首先，通过提取时空流融合特征以充分表达俯视视角下跛行奶牛的运动信息；其次，利用提出的奶牛跛行检测模型Cow-TSM（Cow- Temporal Shift Module）对时空流融合特征图像序列进行特征提取与分类，以检测奶牛是否跛行；最后，对本研究提出方法的有效性进行详细的实验与分析。由于早期跛行的奶牛跛行特征较不明显，并且俯视视角下奶牛跛行特征与侧视视角相比，提取难度较高，因此，本研究的方法主要以检测跛行程度较高的奶牛为主。

1　材料与方法

1.1 数据采集与标注

本研究的实验数据分别于2022年7月和2023年8月在黑龙江省大庆市林甸县晟康牧业牧场采集。采集区域主要有两处：分别是位于挤奶厅中的靠近挤奶区域的挤奶厅通道（图1），以及从牛舍到挤奶厅途中一段宽通道（图2）。摄像头在挤奶厅中的通道采集时距离地面2.8 m，利用1.0 m和0.5 m

长的铝型材搭建采集架，将采集架固定在喷淋管道上；同时在一侧的窗户处放置倾斜45º视角的第2摄像头用来辅助对视频数据进行标注。在位于牛舍去往挤奶厅的宽通道处的摄像头距离地面高度为3.0 m，由于通道较宽，故同时放置3个摄像头进行同步采集，每个摄像头间隔1.0 m，摄像头直接用扎带捆绑在喷淋管道上，并在一侧窗户处水平放置一个摄像头用来辅助对视频数据进行标注。两次采集所使用的深度摄像头为Intel RealSense D435，深度图像的像素分辨率均为848×480，帧率为60 fps。



a. 采集摄像头安装位置



b. 辅助摄像头安装位置

图1 挤奶厅通道采集图

Fig. 1 Collection diagram of milking parlor passage

通道1和通道2两个场景下的奶牛行走视频经过处理后，用于奶牛跛行的数据共包括180段奶牛行走视频流，每段视频分解成100～400帧不等，跛行奶牛和正常奶牛的视频段数比例为1：1。如图3所示，图像经过微调的YOLOv7[17]目标检测模型提取奶牛坐标后，将非奶牛区域像素点置0，而不是将奶牛裁剪，这是为了保留图像序列中奶牛运动时的空间信息。利用YOLOv7模型裁剪奶牛可以避免俯视视角下同一时刻中存在多头奶牛时对目标奶牛的干扰，同时，当奶牛头部发生部分重叠时由于重叠面积较小，对后续实验结果影响较小。

深度视频帧按照Kinetics-400数据集[18]的格式要求进行标注。最后，以每段视频为基准，按照8：2的比例随机划分训练集和测试集。



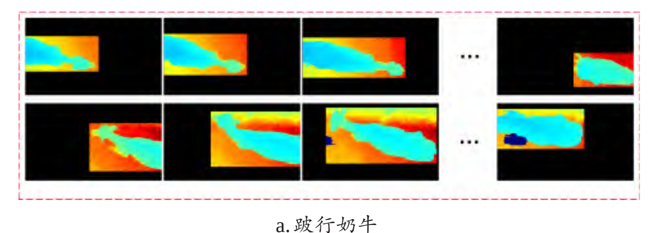
a. 宽通道实际场景示意



b. 采集摄像头安装位置

图2 通往挤奶厅宽通道采集图

Fig. 2 Collection diagram of the wide passage leading to the milking parlor



a. 跛行奶牛

b. 非跛行奶牛

图3 跛行奶牛与非跛行奶牛深度图像序列

Fig. 3 Depth image sequences of lame dairy cows and non-lame dairy cows

1.2 技术路线

跛行奶牛行走过程中，由于患病肢蹄的疼痛会导致躯体在垂直方向上的运动出现不规律性。此外，病蹄的疼痛也会导致行走时较为谨慎，步伐较小，且身体也会出现不同程度的左右摇摆。本研究从上述两方面典型跛行特征开展研究，技术路线图如图4所示。首先提取奶牛时空流融合图像序列，利用YOLOv7目标检测模型检测后的奶牛跛行深度流分为两个分支：分支1使用FlowNet 2.0网络进行

光流图像序列提取，获得时间流特征图像序列；分支2通过深度图相邻帧作差提取深度差值图像，获得空间流特征图像序列。然后将时间流与空间流进行融合，获得时空流融合图像序列。利用时空流融合图像序列作为输入，构建奶牛跛行检测模型Cow-TSM，并进行奶牛跛行检测。

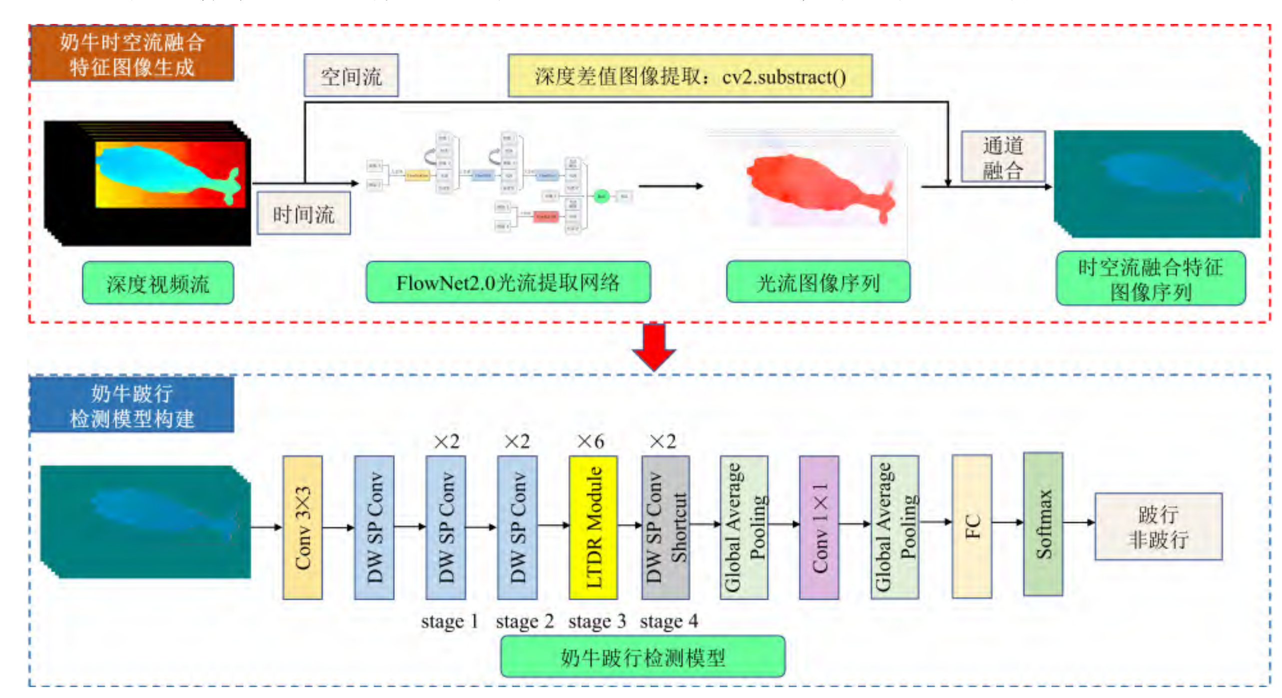


图4　时空流特征融合方法技术路线

Fig. 4 Technical route of spatiotemporal flow feature fusion method

1.3 时空流融合图像序列生成

1.3.1　时间流特征提取

光流能够有效地提取出二维图像中奶牛的平面位移信息。跛行奶牛在运动时，由于步态的不平衡性会导致步长较小，整体运动缓慢，像素点运动位移距离会和正常奶牛出现一定的区别。同时，跛行奶牛在行走时左右摆动的幅度也会随着时间推移产生一定的变化，这些位移变化特征都可以通过光流进行有效提取，因此，本研究从二维奶牛图像中的光流信息中提取出时间流特征。

基于传统方法的光流估计已经较为成熟，常见的稠密光流提取算法包括Farneback算法[19]和Horn-Schunck算法[20]等；稀疏光流提取算法包括Lucas-Kanade算法[21]等。但这些方法都有恒定的假设：图像亮度不变且物体运动缓慢。在实际生产环境中，这些因素都是不可控的，因此传统方法的鲁棒性受到极大的限制。此外，为了提取更多的特征，提取稠密光流是可行的，然而稠密光流的提取计算较为复杂，时效性较差。基于深度学习的光流提取算法对于图像特征的处理更加灵活，在假设提出条件更少的前提下，CNN的层级架构能够提取更抽象、更深入和多尺度的特征，并且计算速度更快[22]。

本研究选用经典的FlowNet 2.0网络[23]进行光流提取。FlowNet 2.0总体结构如图5所示。主体网络利用FlowNetS、FlowNetCorr和FlowNet-SD进行网络堆叠并构成双分支，上面的分支用来堆积成大位移网络提取大位移特征，下面的分支组成小位移网络进行小位移的预测。FlowNetS由CNN卷积模块构成，接收两个RGB图像进行有监督训练。FlowNetCorr与FlowNetS类似，不同的是其首先创建两个相同的网络分支，在网络的高层中的关联层进行计算相关性并把两分支进行合并。FlowNet-SD拥有更大尺寸的输入特征图。在每个子网络后，光流被扭曲并且会和第2张图像进行比较，计算得到的误差在经过其他大位移子网络后，最终输送给融合网络中。将估计的光流、光流的幅度和经过扭曲后的亮度差作为输入，融合网络会进行收缩并尺度扩张，产生最终的光流。

1.3.2　空间流特征提取

跛行奶牛在正常行走时，由于患病肢蹄部位带来的疼痛，会出现不同程度的点头、肢蹄落地位置分布不均和躯体运动不平衡等特征来抵消痛苦。在

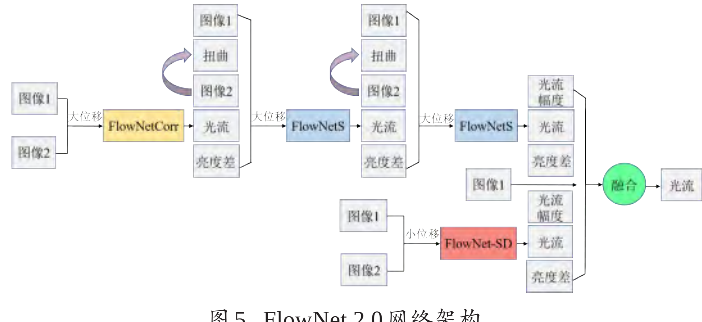


图5 FlowNet 2.0网络架构

Fig. 5 Network structure diagram of FlowNet 2.0

俯视视角下，上述特征主要体现为奶牛运动过程中躯体高低起伏的不规律性。深度图反映了奶牛距相机的距离，为了捕获跛行奶牛在运动时高低起伏的运动特征，将相邻两帧深度图作差，深度的差值反映了相邻帧的奶牛在高度上的步态变化。将相邻两帧深度图像利用OpenCV库中的subtract( )函数进行逐像素减法，在像素相减如果发生溢出时选择取饱和值0。

为了方便后续的通道融合，将深度差值图像进行灰度映射以方便融合。由于深度图像是一种单通道的灰度图像，像素点的数据类型为uint16。为了便于各种算法和神经网络的使用，需要将深度图像进行灰度映射。灰度映射是通过某种映射规则，将原始的灰度像素点根据此种映射规则赋予1个新的灰度值，使整体的像素分布保持不变。在研究中，使用公式（1）作为映射规则，将uint16的深度图像映射为uint8的灰度图像，使每个像素点的范围保持在0～255的区间范围内。

$$newimg = \frac{img - img_{min}}{img_{max} - img_{min}} \times 255 \tag{1}$$

式中：img为原图像像素值；img_{max}为原图像中最大像素点的值；img_{min}为原图像最小像素点的值；$newimg$为映射后的新图像。

1.3.3 时空流融合特征图像生成

奶牛深度差值图像序列体现了奶牛行走过程中的垂直方向上的空间变化特征；奶牛光流图像序列体现了奶牛行走过程中的平面方向上的时间变化轨迹。在研究中，将深度差值图像与光流图像进行通道间融合，将两种模态的数据融合互补，融合方法为利用OpenCV库中的merge( )函数进行通道合并，以此提取奶牛时空流融合特征图像，融合后的部分图像如图6所示。该时空流融合图像拥有奶牛行走过程中的时空特征，再通过后续的模型进行特征提取和建模，来实现有效的奶牛跛行检测。



图6 融合时空流特征模式的奶牛俯视图像

Fig. 6 Top view images of dairy cows fused with spatiotemporal flow feature patterns

1.4 奶牛跛行检测模型构建

基于Cow-TSM的奶牛跛行检测模型结构如图7所示。输入后的主干部分包括4个stage：stage 1和stage 2利用大量深度可分离卷积提取特征的同时减少模型参数；在stage 3使用融合了轻量级时序注意力（Lightweight Temporal Attention, LTA）、时间偏移模块（Temporal Shift Module, TSM）、深度卷积（Depthwise Convolution, DW Conv）、CBAM [24] 注意力机制和重参卷积（Re-parameterization Convolution, REP Conv）的LTDR模块以期利用2D卷积操作实现近似于3D卷积的效果；stage 4使用残差结构的深度可分离卷积。接着经过全局平均池化，1×1卷积调整通道，再进行全局平均池化，全连接和softmax输出视频类别。

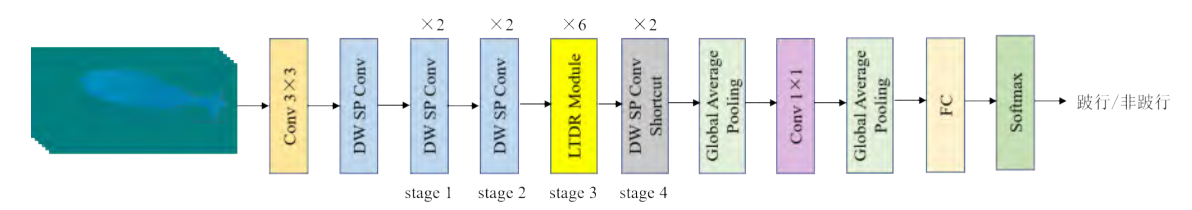


图7 Cow-TSM网络结构

Fig. 7 Network structure of Cow-TSM

模型主体由PP-TSMv2网络构成，该网络是百度飞桨在TSM模块 [25] 之上改进的一种工业落地的2D网络，与TSM相比，有效地改善了原模型的推理精度，推理速度也得到了较大的提升。PP-

TSMv2网络将LTA、TSM模块和带有重参的深度可分离卷积进行组合。本研究在此基础上，在特征提取时添加CBAM注意力机制，如图8所示。首先，LTA模块通过利用全局平均池化和FC层提取全局尺度的时序注意力。其次，将LTA模块的具有全局时序信息的输出经过TSM再进行时序建模。最后，利用深度可分离卷积进行特征提取，深度可分离卷积和重参技术能够有效地降低模型参数量和推理成本，并结合CBAM注意力机制对特征向量进行通道和空间加权，来提升该模块中特征提取的效果。

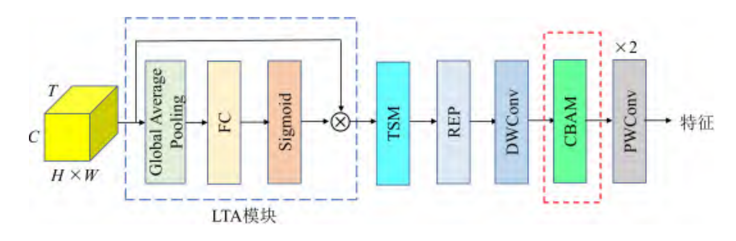


图8　LTDR模块

Fig. 8 LTDR module

1.4.1　时间偏移模块

TSM的核心思想为让特征张量在时间维度的方向上来移动相邻通道来实现相邻帧之间的信息交互，其本身作为一个可插拔的模块可以很方便地插入2D卷积中，而不显著增加计算量和参数，使其兼具2D卷积和3D卷积的优势。图9a为模块输入的原始张量，每一种颜色代表不同时间点的图像帧。接着，在时间维度上将部分通道向下偏移，再将相邻的部分通道向上偏移，原始位置用0补齐，突出的通道截断舍弃，如图9b所示。此时在通道维度上，当前帧包含了前后相邻帧的信息，实现时间维度上的信息交互。在实时推理时，由于下一帧无法被预知，因此只能将上一帧沿着时间维度往下偏移，即由过去向未来，而不能将下一帧沿着时间维度往上偏移，如图9c所示。

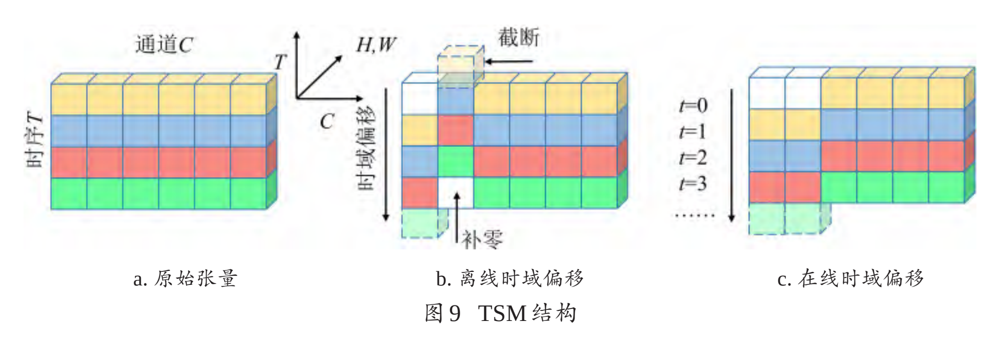


图9　TSM结构

Fig. 9 The structure of TSM

1.4.2　CBAM注意力机制

在提取奶牛时空流融合图像序列后，由于图像的不同通道代表着不同的模态，为了更好地利用图像中的时空流信息，利用通道加权来给不同的模态赋予不同权重是十分关键的。此外，在俯视视角下让模型提取跛行奶牛关键运动特征，往往还需要模型能够注意到关键区域。CBAM注意力模块同时包含了通道和空间两种注意力机制，分别实现表征不同通道的权重，以及提取空间像素间不同位置的关键信息。模块先对输入特征向量先进行通道注意力加权，并于输入特征向量相乘，接着进行空间注意力加权，并与通道加权后的特征向量进行乘积运算。

1.5 评价指标

TP（True Positives）是模型预测的真实值为正的正样本数量；FP（False Positives）是模型预测的真实值为负的正样本数量；FN（False Negatives）是模型预测的真实值为正的负样本数量；TN（True Negatives）是模型预测的真实值为负的负样本的数量。

准确率（Accuracy）是所有预测正确的样本对所有正样本和负样本的总和的比率；*Accuracy*的计算如公式（2）。

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \quad (2)$$

2　结果与分析

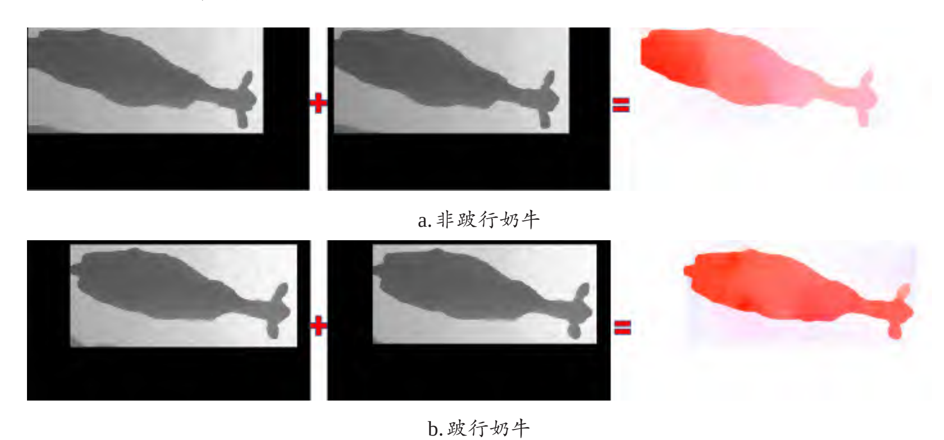
2.1 实验环境与模型参数设置

CPU处理器型号为Intel（R）Xeon（R）CPU E5-2678 v3 @2.50 GHz，系统为Ubuntu 18.04，显卡为RTX 3090，显存大小24 GB，Python版本为3.8，代码编写框架为Pytorch 1.12.0。在训练时，动量系数为0.9，初始学习率设置为0.01，使用余弦退火调整学习率，权重衰减系数为0.000 1，迭代次数设置为150轮次。

2.2 实验结果与分析

2.2.1 光流提取效果

如图10所示，在生成每头奶牛的光流图像时，利用此奶牛的两张相邻灰度图像作为输入，经过微调的FlowNet2.0网络进行预测，输出对应的一帧奶牛光流图像。奶牛光流图像是一种二通道灰度图像，由两个单通道图像拼接得到，分别表征了奶牛在横、纵方向上的位移变化，其中变化程度较大的部分颜色较深。在图10的结果中，将二通道光流图像进行了三通道映射来方便展示。



a.非跛行奶牛

b.跛行奶牛

图10 奶牛灰度图像提取光流可视化

Fig. 10 Optical flow visualization of grayscale image extraction of dairy cows

FlowNet 2.0与传统的稠密光流提取算法进行推理时间的对比，以Farneback算法为例，其提取一张光流图像的平均推理时间平均为0.21 s，FlowNet 2.0约为0.031 s，证明基于FlowNet 2.0的光流提取算法比传统的稠密光流提取算法在时效性上更加优越。

2.2.2 时空流融合特征有效性分析

Cow-TSM模型训练损失曲线如图11所示，模型训练轮数设置为150，从损失曲线图中可以看出，模型在120轮次左右收敛。

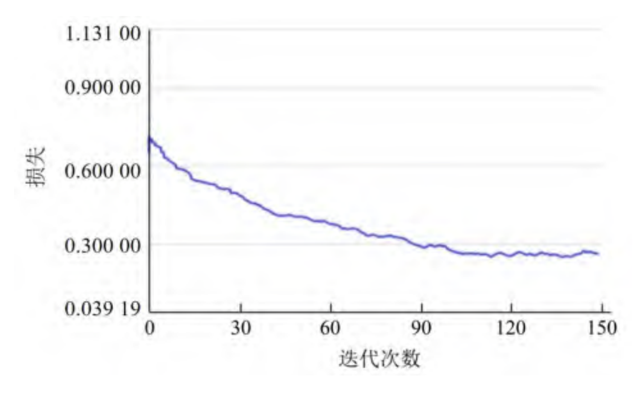


图11 奶牛跛行检测研究Cow-TSM模型训练损失曲线

Fig. 11 Cow-TSM model training loss curve for dairy cow lameness detection

为了验证时空流融合特征的有效性，分别将奶牛光流图像序列、深度图像序列、深度差值图像序列和时空流融合特征图像序列作为Cow-TSM模型的输入，进行实验比较，结果如表1所示。当模型的输入为深度图像序列时，比光流图像序列的准确度高7.5%，这是因为在俯视视角下跛行奶牛的主要特征如点头、四肢运动不规律比较容易被深度图像捕获，因此垂直维度上的不规则运动特征比平面维度上的位移特征能更突出地区分出奶牛是否跛行。当模型的输入为深度差值图像序列时，准确度比深度图像序列高出1.3个百分点，深度差值表征了奶牛的躯体表面在高度方向上的相对位移，这些位移包括了奶牛头部和背部关键部位如髋关节的高度变化。与原始深度图像相比，作差能够减少因摄像头高度不一致带来的误检问题，拥有更强的鲁棒性。把经过光流图像和深度差值图像融合后的时空流融合特征图像序列作为输入时，模型的预测准确度达到88.7%，表现最佳，证明了时间流特征和空间流特征融合的有效性。

表1 奶牛跛行检测研究不同输入图像序列下的模型预测表现

Table 1 Comparison of prediction performances of models with different input images sequences for dairy cow lameness detection research

模型输入序列	*Accuracy*/%	*Precision*/%	*Recall*/%
光流图像序列	76.7	75.3	77.8
深度图像序列	83.4	81.3	84.6
深度差值图像序列	84.6	83.2	85.1
时空流融合特征图像序列	88.7	87.3	89.2

为了保证模型的鲁棒性，在模型训练时使用5折交叉验证，结果如图12所示。将数据集分成互斥的5份，其中的4份作为训练集进行模型的训练，剩下的1份作为测试集。经过一次训练后，将前4份依次作为测试集，进行5次训练，然后对5次训练结果取平均，得到最终的模型表现。

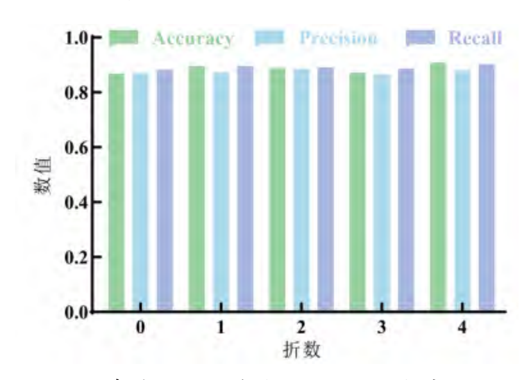


图12　奶牛跛行检测模型Cow-TSM 5折交叉验证

Fig. 12 5 folders cross validation of Cow-TSM model for lameness detection in dairy cows

2.2.3　不同注意力机制对比

为了验证本研究所使用的CBAM注意力对所提出模型的有效性，将其分别与通道注意力（Squeeze-and-Excitation, SE）模块、坐标注意力（Coordinate Attention, CA）模块和卷积核注意力（Selective Kernel, SK）模块进行比较，结果如表2所示，可以得出，融合CBAM模块的模型预测精度为88.7%，CA模块的精度为87.1%，SK模块的精度为86.4%，SE模块为PP-TSMv2模型原有的注意力机制，其预测精度为87.6%。综上可以得出，CBAM注意力机制在本研究中表现最佳。这是因为，模型的输入图像序列由两种模态的图像按通道拼接，因此CBAM模块的通道加权对于模型在本研究数据集上的表现至关重要。此外，CBAM模块中的空间注意能够有效地捕捉奶牛表面关键部位的起伏信息，如奶牛头颈部和四肢与背部连接处。因此，Cow-TSM模型更适合于检测奶牛跛行。

表2　奶牛跛行检测研究中不同注意力机制下的模型预测表现

Table 2 Comparison of prediction performance of models with different attention mechanisms for lameness detection in dairy cows

模型	PP-TSMv2+CA	PP-TSMv2+SE	PP-TSMv2+SK	Cow-TSM
Accuracy/%	87.1	87.6	86.4	88.7

2.2.4　不同视频动作分类网络对比

为了验证本研究所构建的跛行检测模型的有效性，本研究还与TSM模型、PP-TSM模型、基于快慢支路的SlowFast模型和基于Transfomer架构的TimesFormer模型在准确度、参数量和推理时间上进行了对比，由于输入为图像帧序列，不包括视频解码所消耗的时间，因此本实验的推理时间仅为模型的推理时间。

表3展示了Cow-TSM模型与上述视频动作分类网络的对比，由表3可知，Cow-TSM模型在精度上优于主流的基于3D卷积的SlowFast模型，提升1.6个百分点，这是因为使用时序位移模块能在2D卷积操作的基础上实现接近3D卷积的性能，并且更加注重通道与空间加权能力。其次，与其他网络相比，性能也能保持在最佳。通过结合轻量级全局注意力模块LTA与TSM模块，不仅利用到了TSM模块的时序信息，也可以捕获全局时序信息的建模能力，保持优秀的特征提取能力。Cow-TSM模型所消耗的推理时间和占用的参数量均为最低，分别为0.046 s和22 M。Cow-TSM的骨干网络以轻量级卷积神经网络PP-LCNetv2为基础构建，通过使用深度可分离卷积和重参技术使得模型的复杂度大大降低。综上所述，在保证准确率的前提下充分考虑推理速度与模型参数占比，本研究提出的模型十分适合部署在算力受限的边缘设备上。

表3　奶牛跛行检测研究中不同模型预测结果比较

Table 3 Comparison of prediction results of different models for lameness detection in dairy cows

模型	*Accuracy*/%	推理时间/s	参数量/M
TSM	66.7	0.063	141
PP-TSM	84.8	0.096	73
SlowFast	87.1	0.176	200
TimesFormer	85.7	0.933	697
PP-TSMv2	86.6	0.041	20
Cow-TSM	88.7	0.046	22

2.2.5　与现有跛行检测方法对比

为了验证本研究所提出方法的有效性和可行性，与现有国内外的研究方法进行了比较。这些研究中同时包括了侧面视角和俯视视角的奶牛跛行检测方法，对比结果如表4所示。基于侧视视角的跛行检测方法的精度普遍较高，这是因为侧面视角可提取特征较多，如奶牛肢蹄的运动特征，这也是对奶牛跛行检测而言最关键的特征。俯视视角下的精度整体上较为逊色，但是其更适合于在牛场中实际部署。本研究的方法与同为俯视视角下的文献［16］相比，预测精度提升幅度为12.23%，这是因

为文献［16］中使用RGB或深度单一模态进行建模，导致特征提取的能力较弱，并且在俯视视角下RGB模态主要注重于表观纹理信息，对于挖掘跛行相关特征贡献较弱。而本研究通过融合时间流与空间流特征，能够更全面地挖掘奶牛在运动过程中躯体运动对跛行识别时的贡献。与文献［15］相比，本研究的方法的预测精度落后7%。文献［15］方法通过计算机视觉的手段提取奶牛髋关节部位的高度变化，髋关节是奶牛躯干与肢蹄直接连接的关键部位，通过感知该部位的高度变化来表征跛行奶牛在运动过程中的肢蹄运动不规律性。但该方法与本研究方法相比，髋关节部位定位过程较为复杂，而算法准确度十分依赖髋关节的定位准确度，并且在通过希尔伯特变换分析髋关节运动曲线的相位时自动化程度较弱。本研究方法的预处理较少，利用目标检测模型定位到奶牛躯体后，再通过深度学习模型提取整体时空流特征，不会因为关键部位提取效果较差时导致模型检测能力的下降。此外，该文献中的实验样本数目较少，本研究的实验数据集的规模是其7.8倍，并且来自于不同的通道，环境更为复杂，因此本研究模型的鲁棒性较高，预测精度较为稳定。

表4 本研究提出的融合时空流融合特征的方法与现有方法对比

Table 4 Rusults comparison between the porposed method and other methods

模型	*Accuracy*/%	数据集视频规模/个	检测视角
Li等[14]	97.20	680	侧视
Jiang等[12]	96.61	243	侧视
Arazo等[16]	84.56	869	侧视
Arazo等[16]	76.47	864	俯视
Jabbar等[15]	95.70	23	俯视
融合时空流特征的奶牛跛行检测方法	88.70	180	俯视

综上所述，虽然俯视视角下的奶牛跛行检测方法的精度有待进一步的提升，但此类方法不会受到多头奶牛和栏杆遮挡等问题干扰，对于摄像头的安装环境要求较低，不需要在牧场现有的基础设施之上提出更多的要求，并且易与其他奶牛智能感知任务进行联动，如奶牛体重估计和奶牛体况评分等。

3 结　论

针对俯视视角下的复杂场景导致单一模态特征效果较差的问题，本研究提出了一种基于时空流融合特征的奶牛跛行检测方法，利用光流提取网络提取时间流特征图像，利用深度差值图像提取空间流特征图像，并进行通道融合构建时空流融合图像序列。构建奶牛跛行检测模型Cow-TSM，对时空流融合图像数据集进行训练和测试，设计实验探究不同输入模态图像、不同注意力机制和不同模型对奶牛跛行检测的影响，并与现有跛行检测方法进行了比较。本研究提出的奶牛跛行检测方法检测准确度达到88.7%，模型推理时间为0.046 s，模型大小为22 M。结果证明在俯视视角下，通过提取时空流融合特征进行奶牛跛行检测是有效的。

但本研究所提出的算法仍有一定的局限性，在未来，需要加入更多早期跛行的奶牛样本，并进一步挖掘俯视视角下奶牛跛行运动特征，以完善本研究的识别算法。此外，实验数据缺乏光照昏暗或者夜间环境中的奶牛数据，因此需要针对不同光照条件的数据调整本研究的模型训练策略，以得到更加鲁棒的检测模型。

利益冲突声明：本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献：

[1] ARCHER S C, GREEN M J, HUXLEY J N. Association between milk yield and serial locomotion score assessments in UK dairy cows[J]. Journal of dairy science, 2010, 93(9): 4045-4053.

[2] 张楷, 韩书庆, 程国栋, 等. 基于高斯混合-隐马尔科夫融合算法识别奶牛步态时相[J]. 智慧农业(中英文), 2022(2): 53-63.
ZHANG K, HAN S Q, CHENG G D, et al. Gait phase recognition of dairy cows based on Gaussian Mixture model and Hidden Markov model[J]. Smart agriculture, 2022(2): 53-63.

[3] DE MOL R M, ANDRÉ G, BLEUMER E J, et al. Applicability of day-to-day variation in behavior for the automated detection of lameness in dairy cows[J]. Journal of dairy science, 2013, 96(6): 3703-3712.

[4] 李小杉, 杨丰利. 奶牛肢蹄病对繁殖性能的影响[J]. 中国畜牧兽医, 2014, 41(5): 248-251.
LI X S, YANG F L. Effect of lameness on reproductive performance in dairy cows[J]. China animal husbandry & veterinary medicine, 2014, 41(5): 248-251.

[5] CHA E, HERTL J A, BAR D, et al. The cost of different types of lameness in dairy cows calculated by dynamic programming[J]. Preventive veterinary medicine, 2010, 97(1): 1-8.

[6] LEACH K A, TISDALL D A, BELL N J, et al. The effects of early treatment for hindlimb lameness in dairy cows on four commercial UK farms[J]. The veterinary journal, 2012, 193(3): 626-632.

[7] PEZZUOLO A, GUARINO M, SARTORI L, et al. A fea-

sibility study on the use of a structured light depth-camera for three-dimensional body measurements of dairy cows in free-stall barns[J]. Sensors (basel), 2018, 18(2): ID E673.
[8] ZHENG Z Y, ZHANG X Q, QIN L F, et al. Cows' legs tracking and lameness detection in dairy cattle using video analysis and Siamese neural networks[J]. Computers and electronics in agriculture, 2023, 205: ID 107618.
[9] LI Q, CHU M Y, KANG X, et al. Temporal aggregation network using micromotion features for early lameness recognition in dairy cows[J]. Computers and electronics in agriculture, 2023, 204: ID 107562.
[10] POURSABERI A, BAHR C, PLUK A, et al. Real-time automatic lameness detection based on back posture extraction in dairy cattle: Shape analysis of cow with image processing techniques[J]. Computers and electronics in agriculture, 2010, 74(1): 110-119.
[11] WU D H, WU Q, YIN X Q, et al. Lameness detection of dairy cows based on the YOLOv3 deep learning algorithm and a relative step size characteristic vector[J]. Biosystems engineering, 2020, 189: 150-163.
[12] JIANG B, SONG H B, WANG H, et al. Dairy cow lameness detection using a back curvature feature[J]. Computers and electronics in agriculture, 2022, 194: ID 106729.
[13] KANG X, LI S D, LI Q, et al. Dimension-reduced spatiotemporal network for lameness detection in dairy cows[J]. Computers and electronics in agriculture, 2022, 197: ID 106922.
[14] LI Z Y, ZHANG Q R, LYU S C, et al. Fusion of RGB, optical flow and skeleton features for the detection of lameness in dairy cows[J]. Biosystems engineering, 2022, 218: 62-77.
[15] ABDUL JABBAR K, HANSEN M F, SMITH M L, et al. Early and non-intrusive lameness detection in dairy cows using 3-Dimensional video[J]. Biosystems engineering, 2017, 153: 63-69.
[16] ARAZO E, ALY R, MCGUINNESS K. Segmentation enhanced lameness detection in dairy cows from RGB and depth video[EB/OL]. arXiv: 2206.04449, 2022.
[17] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]// 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2023: 7464-7475.
[18] KAY W, CARREIRA J, SIMONYAN K, et al. The kinetics human action video dataset[EB/OL]. arXiv: 1705.06950, 2017.
[19] FARNEBÄCK G. Two-frame motion estimation based on polynomial expansion[C]// Image Analysis: 13th Scandinavian Conference, SCIA 2003. Berlin, Germany: Springer. 2003: 363-370.
[20] HORN B K P, SCHUNCK B G. Determining optical flow[J]. Artificial intelligence, 1981, 17(1/2/3): 185-203.
[21] LUCAS B D, KANADE T. An iterative image registration technique with an application to stereo vision[C]// IJCAI'81: 7th International Joint Conference on Artificial Intelligence. Vancouver, Canada: ACM, 1981: 674-679.
[22] TU Z G, XIE W, ZHANG D J, et al. A survey of variational and CNN-based optical flow techniques[J]. Signal processing: Image communication, 2019, 72: 9-24.
[23] ILG E, MAYER N, SAIKIA T, et al. FlowNet 2.0: Evolution of optical flow estimation with deep networks[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2017: 2462-2470.
[24] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[M]// Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.
[25] LIN J, GAN C, HAN S. TSM: temporal shift module for efficient video understanding[C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2019: 7083-7093.

Automatic Detection Method of Dairy Cow Lameness from Top-view Based on the Fusion of Spatiotemporal Stream Features

DAI Xin[1], WANG Junhao[1], ZHANG Yi[1,2], WANG Xinjie[1], LI Yanxing[1],
DAI Baisheng[1*], SHEN Weizheng[1*]

（1. *College of Electrical Engineering and Information*, *Northeast Agricultural University*, *Harbin* 150030, *China*;
2. *College of Information Engineering*, *East University of Heilongjiang*, *Harbin* 150086, *China*）

Abstract:

[Objective] The detection of lameness in dairy cows is an important issue that needs to be solved urgently in the process of large-scale dairy farming. Timely detection and effective intervention can reduce the culling rate of young dairy cows, which has important practical significance for increasing the milk production of dairy cows and improving the economic benefits of pastures. Due to the low efficiency and low degree of automation of traditional manual detection and contact sensor detection, the mainstream cow lameness detection method is mainly based on computer vision. The detection perspective of existing computer vision-based cow lameness detection methods is mainly side view, but the side view perspective has limitations that are difficult to eliminate. In the actual detection process, there are problems such as cows blocking each other and difficulty in deployment. The cow lameness detection method from the top view will not be difficult to use on the farm due to occlusion problems. The aim is to solve the occlusion problem under the

side view.

[Methods] In order to fully explore the movement undulations of the trunk of the cow and the movement information in the time dimension during the walking process of the cow, a cow lameness detection method was proposed from a top view based on fused spatiotemporal flow features. By analyzing the height changes of the lame cow in the depth video stream during movement, a spatial stream feature image sequence was constructed. By analyzing the instantaneous speed of the lame cow's body moving forward and swaying left and right when walking, optical flow was used to capture the instantaneous speed of the cow's movement, and a time flow characteristic image sequence was constructed. The spatial flow and time flow features were combined to construct a fused spatiotemporal flow feature image sequence. Different from traditional image classification tasks, the image sequence of cows walking includes features in both time and space dimensions. There would be a certain distinction between lame cows and non-lame cows due to their related postures and walking speeds when walking, so using video information analysis was feasible to characterize lameness as a behavior. The video action classification network could effectively model the spatiotemporal information in the input image sequence and output the corresponding category in the predicted result. The attention module Convolutional Block Attention Module (CBAM) was used to improve the PP-TSMv2 video action classification network and build the Cow-TSM cow lameness detection model. The CBAM module could perform channel weighting on different modes of cows, while paying attention to the weights between pixels to improve the model's feature extraction capabilities. Finally, cow lameness experiments were conducted on different modalities, different attention mechanisms, different video action classification networks and comparison of existing methods. The data was used for cow lameness included a total of 180 video streams of cows walking. Each video was decomposed into 100–400 frames. The ratio of the number of video segments of lame cows and normal cows was 1:1. For the feature extraction of cow lameness from the top view, RGB images had less extractable information, so this work mainly used depth video streams.

[Results and Discussions] In this study, a total of 180 segments of cow image sequence data were acquired and processed, including 90 lame cows and 90 non-lame cows with a 1:1 ratio of video segments, and the prediction accuracy of automatic detection method for dairy cow lameness based on fusion of spatiotemporal stream features reaches 88.7%, the model size was 22 M, and the offline inference time was 0.046 s. The prediction accuracy of the common mainstream video action classification models TSM, PP-TSM, SlowFast and TimesFormer models on the data set of automatic detection method for dairy cow lameness based on fusion of spatiotemporal stream features reached 66.7%, 84.8%, 87.1% and 85.7%, respectively. The comprehensive performance of the improved Cow-TSM model in this paper was the most. At the same time, the recognition accuracy of the fused spatiotemporal flow feature image was improved by 12% and 4.1%, respectively, compared with the temporal mode and spatial mode, which proved the effectiveness of spatiotemporal flow fusion in this method. By conducting ablation experiments on different attention mechanisms of SE, SK, CA and CBAM, it was proved that the CBAM attention mechanism used has the best effect on the data of automatic detection method for dairy cow lameness based on fusion of spatiotemporal stream features. The channel attention in CBAM had a better effect on fused spatiotemporal flow data, and the spatial attention could also focus on the key spatial information in cow images. Finally, comparisons were made with existing lameness detection methods, including different methods from side view and top view. Compared with existing methods in the side-view perspective, the prediction accuracy of automatic detection method for dairy cow lameness based on fusion of spatiotemporal stream features was slightly lower, because the side-view perspective had more effective cow lameness characteristics. Compared with the method from the top view, a novel fused spatiotemporal flow feature detection method with better performance and practicability was proposed.

[Conclusions] This method can avoid the occlusion problem of detecting lame cows from the side view, and at the same time improves the prediction accuracy of the detection method from the top view. It is of great significance for reducing the incidence of lameness in cows and improving the economic benefits of the pasture, and meets the needs of large-scale construction of the pasture.

Key words: dairy cow lameness detection; spatiotemporal fusion; video action classification; depth image; attention mechanism; TSM

Foundation items: National Natural Science Foundation of China (32072788); Key Research and Development Program of Heilongjiang Province (2022ZX01A24); National Key Research and Development Program of China (2023YFD2000700); Project Supported by Scientific Research Platform of East University of Heilongjiang (PTZCXM2404)

Biography: DAI Xin, E-mail: daixin@neau.edu.cn

***Corresponding author:** 1. DAI Baisheng, E-mail:bsdai@neau.edu.cn; 2. SHEN Weizheng, E-mail: wzshen@neau.edu.cn